Sensor Node Localization using SIFT Algorithm

Daniela Fuiorea*, Vasile Gui*, Dan Pescaru**, Petronela Paraschiv*,
Istin Codruta**, Daniel Curiac**, Constantin Volosencu**
* Faculty of Electonic and Communication, ** Faculty of Automation and Computers
"Politehnica" University of Timisoara
V. Parvan No. 2, 300223 Timisoara, Romania
ROMANIA

Abstract: - Sensor localization is required by most Wireless Sensor Networks applications. Considering application for video surveillance, localization includes not only spatial coordination but also cameras direction and video-field overlap estimation. This paper presents a novel technique for localization in a Video-based Wireless Sensor Network using image registration that involves SIFT algorithm for automatic features selection. Experimental results show the estimation accuracy and time efficiency comparing with manual solution.

Key-Words: - wireless sensor networks, topology, localization, deployment, image processing, SIFT

1 Introduction

Wireless Sensor Networks (WSN) consists of many small, simple, cheap sensor nodes that cooperatively monitor physical conditions, such as temperature, sound, vibration or pressure. The information collected is then delivered to the other nodes over the wireless link. After having been distributed randomly in a given region a very first step consists in self-localization. Precise GPS-based solutions are not feasible due to expensive and power consuming additional hardware involved [1]. Indeed, several other methods were proposed [2][3]. In this paper, discussions are around a novel technique of node localization in Video-based Wireless Sensor Networks. Video is an important medium for the observation of a variety of phenomena in the physical world. For example, in the cameras can be used to monitor different activities, evaluate land erosion, and observe a variety of animal species. Due to the fact that cameras from a Video-based WSN can exist in a great number, the pictures gathered from it could contain images with a common field of view, images that are taken from different position and angles. A reconstruction of the scene would imply combining these images in order to give us a panoramic view over environment. Therefore, an image registration task is involved. However, information gathered by this task will add to basic node localization useful information like cameras' direction and Field of View.

Image Registration [4] is a basic image processing operation employed by many computer vision applications used in areas such as remote sensing, biomedical imaging, surveillance and robotics. Its goal is to overlay two or more images of the same scene taken at different times, from different viewpoints, and/or by different cameras. In order to register two images, a transformation must be found so that each point in one reference image can be mapped to a point in the second image. The geometric transformation in image registration is the similarity transform, consisting of rotation, translation and scaling. The model is defined by the equations:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} s & 1 \\ 1 & s \end{bmatrix} \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \qquad (1)$$

relating the old pixel coordinates (x,y) to the new ones. The four parameters of the transformation can be unambiguously determined from the correspondence of two pairs of points. However, in most of the cases, the number of the points available for estimating the transformation parameters s, φ, t_x and t_y is higher.

The correspondence between the features can be classified in two categories: feature-based and region-based. Due to the fact that the region-based registration is prone to errors generated by segmentation and different color sensitivities of the cameras, image point features can be used instead. This approach has been shown to be more robust with viewpoint, scale and illumination changes, and occlusion. Feature selection can be carried out manually or automatically. While first approach involves a human operator, the second one uses a SIFT derived algorithm [5]. Unfortunately, even is more precise, first approach is to slow in case of a real world application involving thousands of nodes.

The approach investigated in this paper regards to the automatic feature selection used in localization of wireless video sensors. The rest of the paper is organizes as following. Section 2 provides an overview of existing localization techniques. Nest section describes registration using the SIFT algorithm. Section 4 presents evaluation results on real world images. Finally, conclusions of this work are presented in the last section.

2 WSN Localization Techniques

As presented, sensor nodes of a WSN are usually deployed in a random, ad-hoc manner. Often they are scattered from an airplane or moving vehicle on an unplanned infrastructure. The key problem of estimating spatial-coordinates of network nodes is referred to as topology extraction or localization.

A simple solution uses GPS at the level of each node. However, this can work only outdoors and, as the receiver is expensive, large and power consuming, it is not suitable for the construction of small, cheap and energy efficient sensor nodes. Indeed, many localization methods estimate the locations of sensors by using knowledge of the absolute positions of only a few GPS-based sensors based on inter-sensor distance measurements [6].

2.1 Distance Measurement Algorithms

Distance measurement algorithms are used to estimate relative position for network nodes. A coarse classification of these algorithms contains three main categories: angle-of-arrival measurements, distance related measurements and RSS profiling measurements.

The angle-of-arrival measurement techniques can use the receiver antenna's amplitude response or they can consider the receiver antenna's phase response. The accuracy of these techniques is limited by the directivity of the antenna, by shadowing and by multi-path reflections. A multi-path component may appear as a signal arriving from an entirely different direction and can lead to very large errors in AOA measurements. Multi-path problems in AOA measurements can be addressed by using the maximum likelihood (ML) algorithms. Various ML algorithms were developed. The best-known examples are Multiple Signal Classification (*MUSIC*) [7] and Conjugate Estimation of Signal Parameters by Rotational Invariance Techniques (*C-ESPRIT*) [8].

Distance related measurements are based on propagation time or radio signal strength. In the first case one possibility is to estimate distances between neighboring sensors using time-of-arrival. It represents the propagation time of a signal between the transmitter and the receiver. Therefore it requires the local time at the transmitter and the local time at the receiver to be accurately synchronized. This disadvantage makes time-to-arrival time measurements a less attractive [9]. An improved technique is called roundtrip propagation time. It measures the difference between the time when a sensor sends a signal and the time when the answer is received at the original sensor. Since the same clock is used to compute the time synchronization problem is avoided. However, a major error source in roundtrip propagation time measurements is the delay required for handling the signal in the second sensor [10]. Another interesting approach to distance measurements is the lighthouse approach [11] which derives the distance between an optical receiver and a transmitter of a parallel rotating optical beam by measuring the time duration that the receiver dwells in the beam. Weak points of this approach are the cost of optical sensor and requirements of a direct line-of-sight between the optical receiver and the transmitter.

Another category of distance related measurement techniques estimates the distances between neighboring sensors from the received signal strength. They are attractive because they require no additional hardware. However, depending of the deployment environment the propagation of a signal is affected by reflection, diffraction and scattering [12].

As concerning *RSS* profiling-based localization, it works by constructing a form of map of the signal strength behavior in the coverage area. The map is obtained either offline by a priori measurements or online using sniffing devices [13] deployed at known locations. They have been mainly used for location estimation in *WLANs*, but they would appear to be attractive also for wireless sensor networks. The model is stored in a central location. By referring to the *RSS* model, a non-anchor node can estimate its location using the *RSS* measurements from anchors.

2.1 Localization in Context of Video-based Wireless Sensor Networks

In context of video-based Wireless Sensor Networks the localization problem is more complex. In addition to nodes coordinates the information regarding topology includes the angles between cameras and video fields overlapping. Indeed, these extra parameters are essential for most video surveillance applications.

In [14] we propose a novel solution for node localization and video-filed overlap estimation. It starts from video images acquired from different nodes and computes video fields superposition with the help of a central server (could be a PC or a notebook). Then it computes parameters like coordinates translation, rotation angle and scaling factor and diffuses the extracted information into the entire network. Node localization is based on estimation of these parameters between each pair of neighbor nodes. Indeed, it involves image registration applied against static images gathered quasi-simultaneous from entire network. In order to accomplish the registration task for a pair of corresponding images, an important step is the feature selection. Considering the features detected, a transformation is found and each point in one image (node image) is mapped to a point in the second (neighbor node image). In the previous work [14][15], the feature selection was carried out manually and was based on detecting corners by a simply click of the mouse. For this, we need a human operator to interact with the mouse in choosing the corresponding pairs of feature points in the images.

The proposed algorithm was still affected by small errors (1 to 5 pixels) specific to correspondence point setting procedure. To deal with these errors a post-processing step was required and *chamfer-matching* post-processing was considered [16]. However, the main drawback of the method consists in large amount of time involved. The work presented here is motivated by the need of an fast automatic extraction of corresponding points as the starting point for registration.

3 Localization Based on Registration using Automatic Feature Selection for Parameter Estimation

The point mapping technique is a primary approach taken to register two images that type of misalignment is unknown. The general method consists of three steps. In the first step features in the images are computed. The second step is identifying feature correspondences in pairs of images. And the last step is estimating parameters of geometrical transforms optimally mapping features between pairs of images.

Feature selection can be carried out in two ways. We can select feature correspondences manually, using only the button of the mouse or we can select automatically using an algorithm for feature detection.

The automatic selection is based on a feature selection algorithm named SIFT [5]. SIFT is coming from Scale-Invariant Feature Transform which is an algorithm in computer vision to detect and describe local features in images. These local features are invariant to image scale and rotation. They are also robust to changes in illumination, noise, occlusion and minor changes in viewpoint. In addition to these properties, the local features given by SIFT are distinctive, easy to extract, allow for correct object identification with low probability of mismatch and are easy to match against a (large) database of local features and they can be used for matching. In order to generate a set of image features four steps of computation are used:

- Scale-space extrema detection
- Key-point localization
- Orientation assignment
- The local image descriptor

An important aspect of this approach is that it generates large numbers of features that densely cover the image over the full range of scales and locations.

After one of these steps is accomplished, the next step based on a robust estimation method gives the best estimate of the parameters of geometrical transform, mapping features between pairs of images.

4 Testing Results

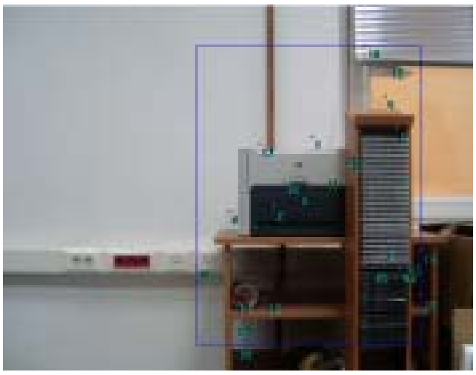
In order to test the performances of the automatic selection, we used image pairs containing a common field of view, obtained for different camera positions and orientations. As test bed we use a five TRENDNET IP-400W wireless camera nodes network.

The real world applications could be deployed in various environments. However, we identify three common situations for most applications. First is represented by nature scenes, like forests or desert; the second is city landscapes, like squares or intersections; and the last is indoor, like rooms or passage. Due to this fact, the experiments were realized with images considering these environments. Consequently, the images were split in three categories: forest, office and urban landscape. An example is given in Figure 1.

For each image, the selection was made automatic and manual for different number of feature points. Each method is tested on 15 images (5 images x 3 categories).

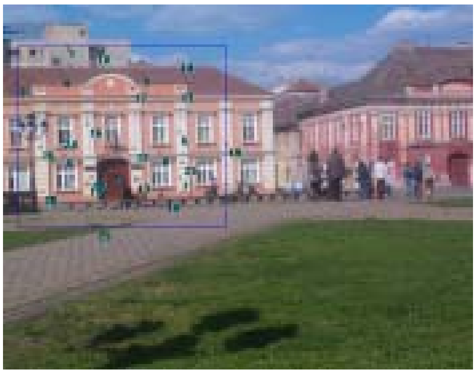
(a)

(b)

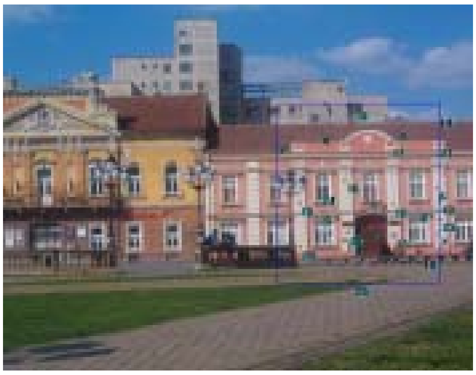


(c)

(d)



(e)



(f)

Fig. 1. Images from different environments:
(a), (b) Forest scene; (c), (d) Office scene; (e), (f) Urban scene

The parameters of the similarity transform (s, φ, t_x and t_y), which are resulting from using manual selection, are considered as the ideal ones.

All comparisons use a relative error computed with the equation:

$$err = \frac{(x-y)}{\dim} * 100, \qquad (2)$$

where x is the ideal value, obtained from using the manual selection, y is the value obtained from using the automatic method and $\dim$ is the horizontal dimension of the image. This equation is applied to every parameter of the similarity transform. The resulting values are presented in Table I.

TABLE I
The relative error *err* (in percentage)

Parameters of similarity transform	**Forest**				
	Img1	**Img2**	**Img3**	**Img4**	**Img5**
Translation X	32	1.7	0.07	0.67	0.35
Translation Y	0.36	1.46	0.08	0.29	1.43
Scale	0.00013	0.0013	0.002	0.001	0.0008
Angle	0.00032	0.0015	0.002	0.002	0.0030

Parameters of similarity transform	**Office**				
	Img1	**Img2**	**Img3**	**Img4**	**Img5**
Translation X	1.58	0.39	0.98	1.69	0.19
Translation Y	0.64	3.18	0.08	0.79	6.51
Scale	0.0007	0.004	0.004	0.0001	0.009
Angle	0.0012	0.001	0.006	0.0005	0.005

Parameters of similarity transform	**Urban**				
	Img1	**Img2**	**Img3**	**Img4**	**Img5**
Translation X	0.001	0.70	0.14	0.55	0.17
Translation Y	0.50	0.59	0.19	1.19	0.021
Scale	0.0004	0.0005	0.0006	0.0023	0.001
Angle	0.0001	0.0010	0.0001	0.0010	0.001

Small errors are present for the scale and angle estimate in all three categories. Horizontal and vertical translation estimate presents small errors for pictures with indoor environment and buildings. The worst case is for the horizontal translation estimate, where the worst-case error is about 32%. The best results are obtained for scale and angle on building images. The vertical translation estimate presents a relative highest error, 6.51%, when an indoor image is used.

The highest and the lowest values of the error are presented in Table II, along with the trimmed mean for every parameter of the similarity transform.

ISBN: 978-960-6766-77-0

ISSN 1790-5117

TABLE II
Trimmed Mean

Parameters of similarity transform	**Highest value**	**Trimmed mean**	**Lowest value**
Translation X	32%	0.71%	0.001%
Translation Y	6.51%	0.83%	0.021%
Scale	0.009%	0.0014%	0.0001%
Angle	0.006%	0.0015%	0.0001%

A trimmed mean is calculated by discarding a certain percentage of the lowest and the highest values of the relative error followed by computing the mean of the remaining values. The trimmed mean is a family of measures. The $\alpha\%$ - trimmed mean of N values $x_1,...,x_n$ is computing by sorting all the N values, discarding $\alpha\%$ of the lowest and $\alpha\%$ of the highest values, and computing the mean of the remaining values. In our case, a trimmed mean $\alpha = 7\%$ is computed for $N = 15$.

TABLE III
The execution time

Manual			**Automatic**		
25 pts			**25 pts**		
Parc	**Office**	**Urban**	**Parc**	**Office**	**Urban**
4’10”	3’10”	2’10”	15”	12”	9”
15 pts			**15 pts**		
3’5”	1’30”	1’30”	14”	11”	8”

Table III presents the time measurement for each category in the case of 25 points and 15 points. In a real WNS the number of gather images is large. Based on the measurements from the Table III, we can estimate the time involved by a real application. Suppose we gather 500 pairs of images. When the number of considered points is 25 and the selection is manually, the registration process will take about 4’10” x 500 pairs of images resulting in 35 hours. When the feature selection is automatic, the estimated time for registration process will be 15”x 500 pairs of images that means around 2 hours. Even considering the best case of city landscapes, for manual selection the localization time is unacceptable (around 18 hours) while the automatic selection takes just 1 hour.

5 Conclusion

This paper proposes a localization algorithm based on registration applied on images gathered from network nodes when using an automatic features selection using SIFT algorithm. In addition to spatial localization we estimate also the video-field overlap between each pair of camera-nodes. Test results demonstrate the benefit of this technique in terms of execution time while loosing in precision is acceptable.

References:

[1] A. H. Sayed, A. Tarighat, and N. Khajehnouri, Network-based wireless location: challenges faced in developing techniques for accurate wireless location information, *IEEE Signal Processing Magazine*, Vol. 22, No. 4, 2005, pp. 24–40.

[2] C. Buschmann, H. Hellbruck, S. Fischer, A. Kroller and S. P. Fekete, Radio Propagation-Aware Distance Estimation Based on Neighborhood Comparison, *in Proc. of the 4th European Conference Wireless Sensor Networks, EWSN 2007*, Delft, Netherlands, January 29-31, 2007, pp. 325-340.

[3] G. Mao, B. Fidan, and B. Anderson, Wireless sensor network localization techniques, *ACM Computer Networks Journal*, Vol. 51, No. 10, July 2007, pp. 2529-2553.

[4] B. Zitova, J. Flusser, “Image registration methods: a survey”, *Image Vision and Computing* 21, Elsevier, 2003, pp. 977-1000.

[5] D. G. Lowe, Distinctive image features from scale-invariant key-points, *International Journal of Computer Vision,* Vol. 60, 2004, pp. 91-110.

[6] S. Guolin, C. Jie, G. Wei, and K. J. R. Liu, Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs, *IEEE Signal Processing Magazine*, Vol. 22, No. 4, 2005, pp. 12–23.

[7] M. Kaveh and A. Bassias, Threshold extension based on a new paradigm for MUSIC-type estimation, *in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, 1990, pp. 2535–2538.

[8] N. Tayem and H. M. Kwon, Conjugate ESPRIT (C-SPRIT), *IEEE Transactions on Antennas and Propagation*, Vol. 52, No. 10, 2004, pp. 2618–2624.

[9] E. Elnahrawy, X. Li and R. P. Martin, The Limits of Localization Using Signal Strength: A Comparative Study, *In Proceedings of the First IEEE International Conference on Sensor and Ad hoc Communications and Networks (SECON 2004)*, Santa Clara, CA, USA, October 2004, pp. 41-52.

[10] D. McCrady, L. Doyle, H. Forstrom, T. Dempsey, and M. Martorana, Mobile ranging using low-accuracy clocks, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 48, No. 6, 2000, pp. 951–958.

[11] K. Romer, The lighthouse location system for smart dust, *in Proceedings of ACM Conference on Mobile Systems, Applications, and Services*, 2003, pp. 15–30.

[12] E. Elnahrawy, X. Li, and R. Martin, The limits of localization using signal strength: a comparative study, *in Proceedings of the First Annual IEEE Conference on Sensor and Ad-hoc Communications and Networks*, 2004, pp. 406–414.

[13] P. Krishnan, A. Krishnakumar, W.-H. Ju, C. Mallows, and S. Gamt, A system for LEASE: location estimation assisted by stationary emitters for indoor RF wireless networks, *in Proceedings of IEEE INFOCOM'04*, Vol. 2, 2004, pp. 1001–1011.

[14] D. Pescaru, D. Fuiorea, V. Gui, C. Toma and G. Muntean, Image-based Node Localization Algorithm for Wireless Video Sensor Networks, *in Proc. of the 6th Information Technology and Telecommunication Conference, IT&T'06*, Ireland, 2006, pp. 139-147.

[15] D. Fuiorea, V. Gui, D. Pescaru and C. Toma, Using Registration Algorithms for Wireless Sensor Network Node Localization, *In Proc. of the 4th IEEE International Symposium on Applied Computational Intelligence and Informatics SACI2007*, May 17-18, 2007, pp. 209-214.

[16] G. Borgefors, "Hierarchical Chamfer Matching: A parametric edge matching algorithm", *IEEE Transaction on Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 6, 1988, pp. 849-865.